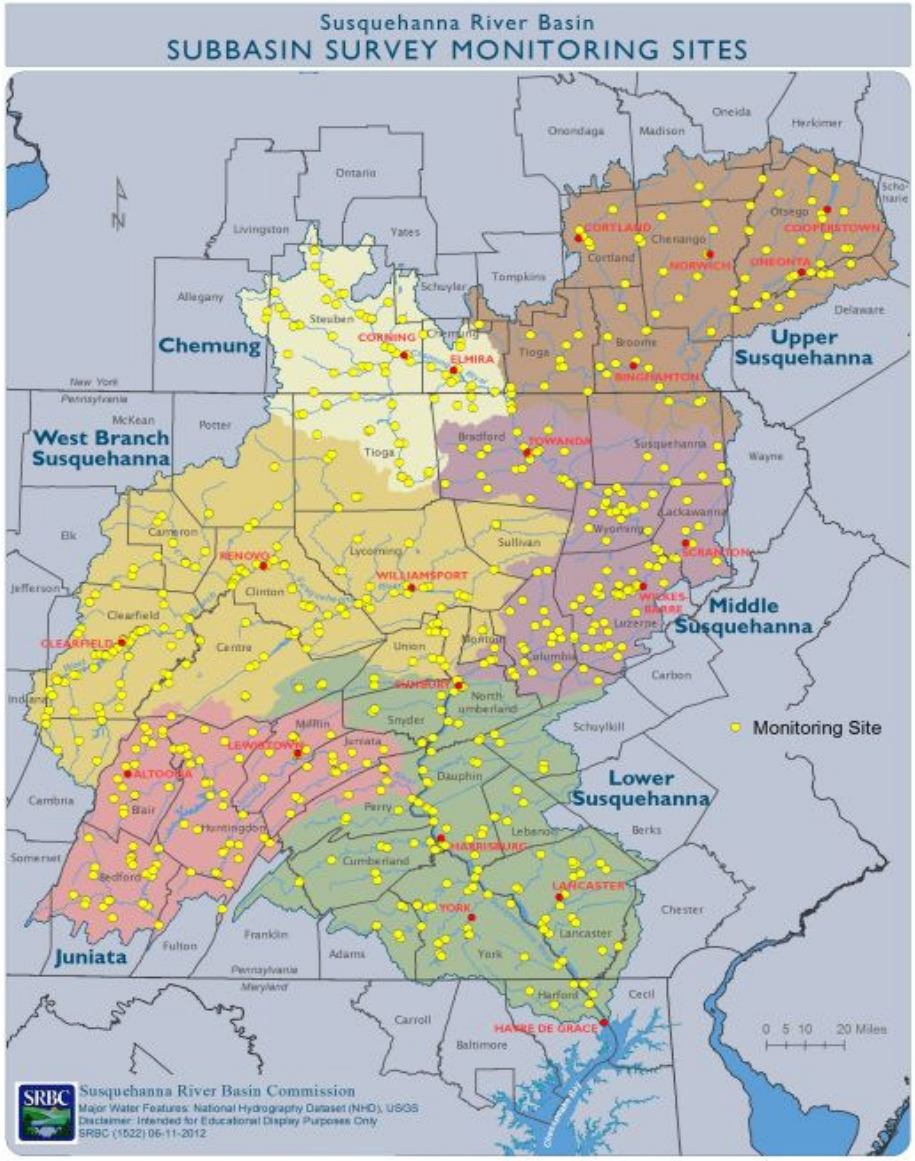
Sampling Design
Targeted vs. Probabilistic

Water Quality Advisory Committee

May 22, 2013

Subbasin Surveys

- A broad brush look at each of the 6 major subbasins on a rotating basis; started in the 1980s.
- Sites chosen primarily by ease of access
  - mouths of large tributaries
  - larger watersheds arbitrarily got more sites
  - bridge crossings



Targeted Site Selection

- Targeted sampling is a very useful and valid type of sample design if project goals include:
  - characterization of regional reference conditions
  - long-term trend monitoring at specific sites
  - bracketing a known source, i.e. discharge point
  - Inferences can ***only*** be applied to sites where data was collected due to biases associated with site selection
    - cannot apply results to entire study area

Randomized Site Selection

- Probabilistic sampling design is useful when project goals include:
  - Aquatic life use attainment in un-biased, statistically sound manner across large areas (i.e. state, watershed)
  - Uniform comparability between datasets and other national datasets
  - **Allows inference from the sample data to the status of the entire study area[1]**
  - Necessary to avoid the biases, even unrecognized ones, inherent in selecting sample sites[1]

[1] Herlihy et al, 2000

- In current Subbasin Survey format we may sample 100 sites in a subbasin and still not make any inferences about the ecological conditions in the subbasin due to the bias in site selection
  - 50% **of the streams we sampled** in the Juniata subbasin were nonimpaired

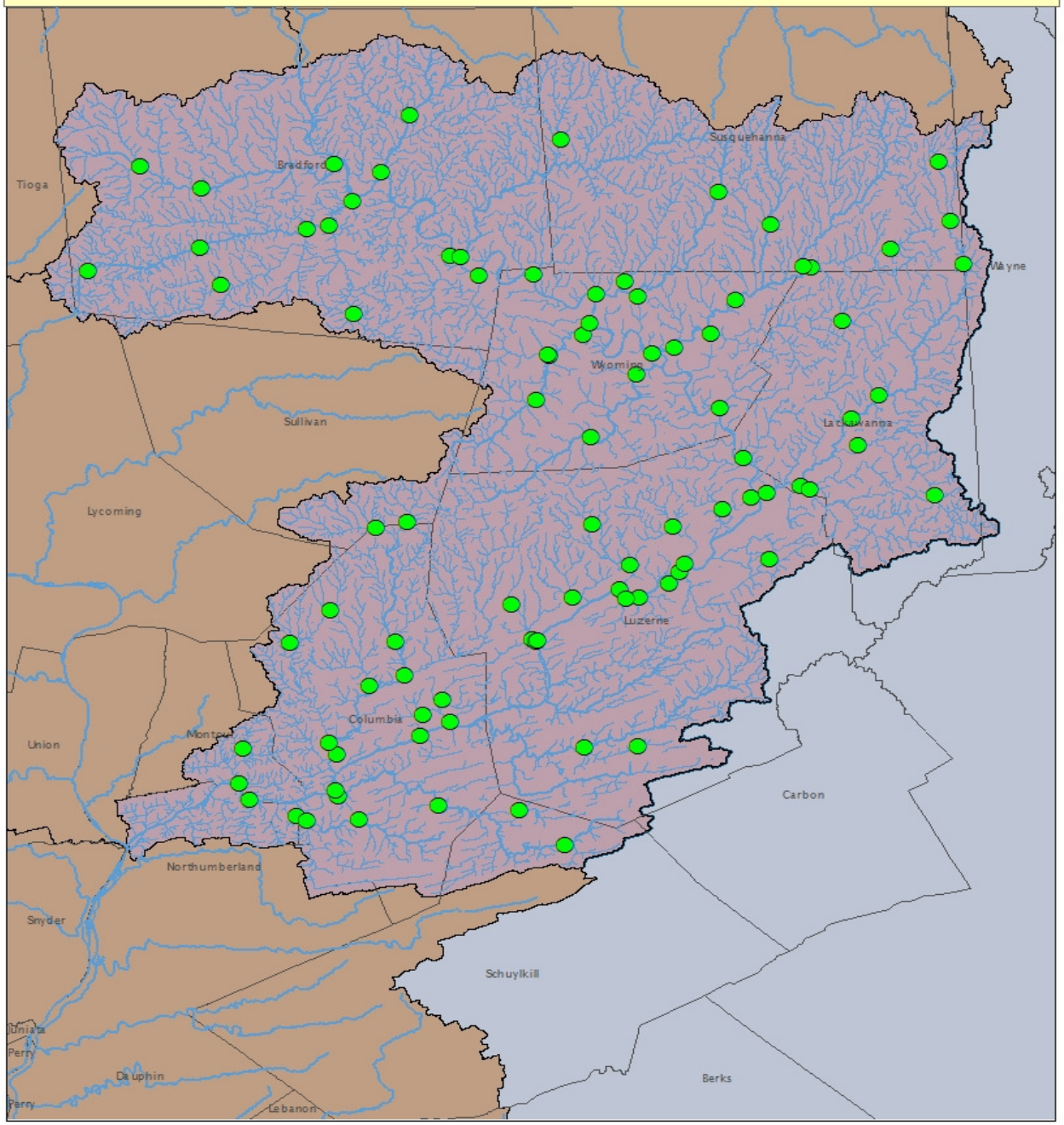
VS.

  - 50% **of the streams** in the Juniata subbasin were nonimpaired

Proposed new sample design

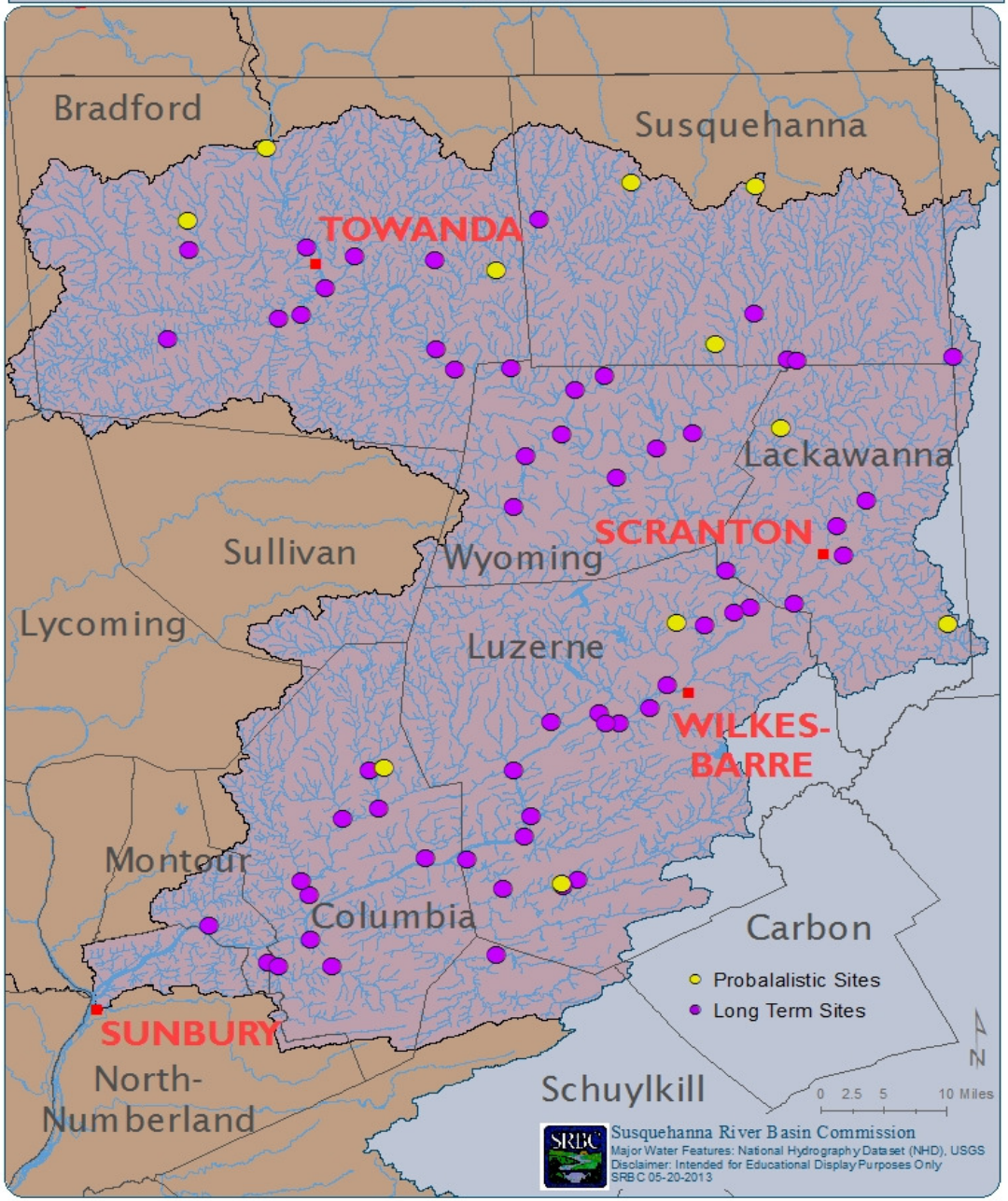
- Middle Subbasin slated for 2014
- Keep some targeted sites from past surveys; major tributaries, reference condition sites
- Pilot the inclusion of a set number of random sites per subbasin
- The remaining sites will be designated for streams of concern to member jurisdictions, unassessed waters, or other specific targeted interests
- Remove mainstem Susquehanna River sites from Subbasin Survey and rely on Large River program to assess those sites

2008 Middle Susquehanna Subbasin Sampling Sites



Susquehanna River Basin

MIDDLE SUBBASIN POSSIBLE 2014 MONITORING SITES

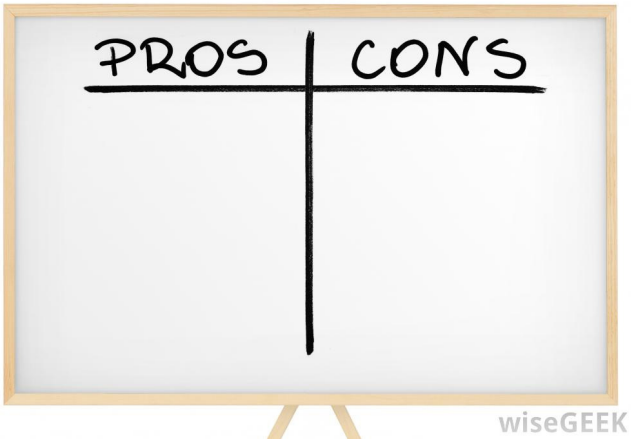


Generalized Random Tessellation Stratified (GRTS) Design

- GIS- based approach with the site selection done using the R stat program
- Allow for simple or more complex designs
  - Equal probability
    - every segment has the same chance of being selected
  - Unequal probability
    - make certain features more likely to be sampled (i.e. stream order, drainage area)
  - Oversample Sites
    - select additional sites to account for unforeseen complications (i.e. dry streams, permission/access issues)
  - Panels for surveys over time
    - i.e. 3 year project in same watershed, new random sites every year
  - Two stage sampling using GRTS at each stage
    - i.e. Randomly select HUC 12 watersheds, then randomly select sites within those watersheds

Drawbacks

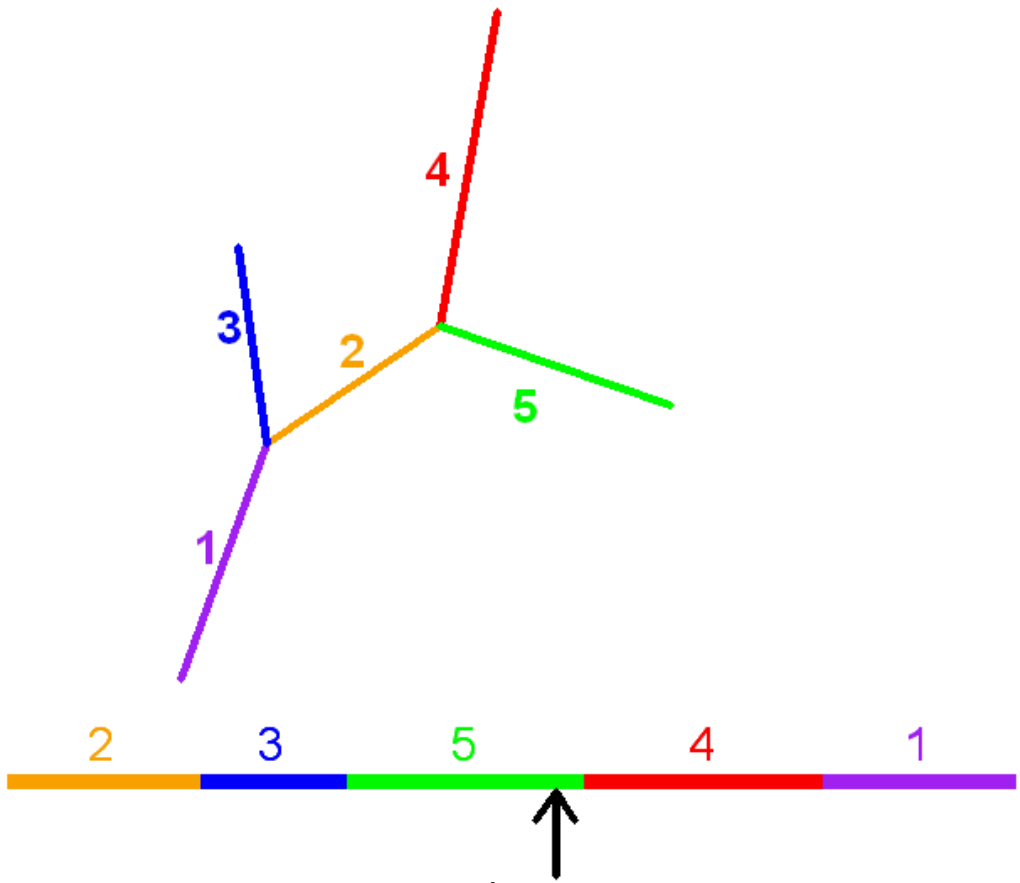
- Sites may not exist, may be dry/intermittent or not be wadeable
- Extra time and staff resources:
  - generate random sites and oversample sites
  - locate and verify sites
  - secure permission, if needed
  - potential added time to reach sites
  - replacing and verifying sites from oversample list
- Can do fewer sites with the same amount of resources



Feedback Questions

Generalized Random Tessellation Stratified (GRTS) Design

Simple example of randomly selecting one site in the below watershed:



- Segment 5 is selected at point 8/9 along its representation on the line
- Calculate sample site location as 8/9 of segment length upstream from node